

Dynamical Variational Autoencoders: A Comprehensive Review

Year: 2020 | Citations: 256 | Authors: Laurent Girin, Simon Leglaive, Xiaoyu Bie, Julien Diard, Thomas Hueber

Abstract

The Variational Autoencoder (VAE) is a powerful deep generative model that is now extensively used to represent high-dimensional complex data via a low-dimensional latent space that is learned in an unsupervised manner. In the original VAE model, input data vectors are processed independently. In the recent years, a series of papers have presented different extensions of the VAE to sequential data, that not only model the latent space, but also model the temporal dependencies within a sequence of data vectors and/or corresponding latent vectors, relying on recurrent neural networks or state space models. In this paper we perform an extensive literature review of these models. Importantly, we introduce and discuss a general class of models called Dynamical Variational Autoencoders (DVAEs) that encompass a large subset of these temporal VAE extensions. Then we present in details seven different instances of DVAE that were recently proposed in the literature, with an effort to homogenize the notations and presentation lines, as well as to relate those models with existing classical temporal models (that are also presented for the sake of completeness). We reimplemented those seven DVAE models and we present the results of an experimental benchmark that we conducted on the speech analysis-resynthesis task (the PyTorch code will be made publicly available). An extensive discussion is presented at the end of the paper, aiming to comment on important issues concerning the DVAE class of models and to describe future research guidelines.